Special Report 16

Circular Statistical Methods: Applications in Spatial and Temporal Performance Analysis

Robert P. Mahan
The University of Georgia

April 1991

United States Army Research Institute
for the Behavioral and Social Sciences

U.S. ARMY RESEARCH INSTITUTE
FOR THE BEHAVIORAL AND SOCIAL SCIENCES

A Field Operating Agency Under the Jurisdiction of the Deputy Chief of Staff for Personnel

EDGAR M. JOHNSON
Technical Director

JON W. BLADES
COL, IN
Commanding

Research accomplished under contract for the Department of the Army

The University of Georgia

Technical review by

David W. Bessemer

NOTICES

SECURITY CLASSIFICATION OF THIS PAGE

REPORT DOCUMENTATION PAGE

Form Approved
OMB No. 0704-0188

1a. REPORT SECURITY CLASSIFICATION
Unclassified

1b. RESTRICTIVE MARKINGS
--

2a. SECURITY CLASSIFICATION AUTHORITY
--

2b. DECLASSIFICATION / DOWNGRADING SCHEDULE
--

3. DISTRIBUTION / AVAILABILITY OF REPORT
Approved for public release; distribution is unlimited.

4. PERFORMING ORGANIZATION REPORT NUMBER(S)
Delivery Order 1910 TCN 90156

5. MONITORING ORGANIZATION REPORT NUMBER(S)
ARI Special Report 16

6a. NAME OF PERFORMING ORGANIZATION
The University of Georgia

6b. OFFICE SYMBOL *(If applicable)*
--

7a. NAME OF MONITORING ORGANIZATION
U.S. Army Research Institute

6c. ADDRESS *(City, State, and ZIP Code)*
Department of Psychology
Athens, GA 30602

7b. ADDRESS *(City, State, and ZIP Code)*
ATTN: PERI-IK
Ft. Knox, KY 40121-5620

8a. NAME OF FUNDING / SPONSORING ORGANIZATION
U.S. Army Research Institute for the Behavioral and Social Sciences

8b. OFFICE SYMBOL *(If applicable)*
PERI-I

9. PROCUREMENT INSTRUMENT IDENTIFICATION NUMBER
DAAL03-86-D-0001

8c. ADDRESS *(City, State, and ZIP Code)*
5001 Eisenhower Avenue
Alexandria, VA 22333-5600

10. SOURCE OF FUNDING NUMBERS

PROGRAM ELEMENT NO.	PROJECT NO.	TASK NO.	WORK UNIT ACCESSION NO.
N/A	N/A	N/A	N/A

11. TITLE *(Include Security Classification)*
Circular Statistical Methods: Applications in Spatial and Temporal Performance Analysis

12. PERSONAL AUTHOR(S)
Mahan, Robert P.

13a. TYPE OF REPORT	13b. TIME COVERED	14. DATE OF REPORT *(Year, Month, Day)*	15. PAGE COUNT
Final	FROM 90/06 TO 90/10	1991, April	53

16. SUPPLEMENTARY NOTATION Task was performed under a Scientific Services Agreement issued by Battelle, Research Triangle Park Office, 200 Park Drive, P.O. Box 12297, Research Triangle Park, NC 27709, and U.S. Army Research Office, P.O. Box 12211, Research (Continued)

17. COSATI CODES

FIELD	GROUP	SUB-GROUP

18. SUBJECT TERMS *(Continue on reverse if necessary and identify by block number)*
Circular scales
Circular statistics
Spatial performance
Temporal performance
Performance measures
Networked simulators

19. ABSTRACT *(Continue on reverse if necessary and identify by block number)*

This report surveys methods used for analyzing data generated from circular scales. The focus is on dimensions of simulation network training and development data that may be inaccessible to linear-based analyses. Several problem areas are identified that may lead to erroneous conclusions and/or loss of information when applying traditional analytic procedures to spatial and temporal performance measures. For example, the arithmetic mean and standard deviation are shown to be inappropriate descriptive measures for circular data. In addition, the use of average absolute deviations to measure directional errors of judgment can lead to loss of directional information. Finally, the usual methods of statistical inference are shown to fail in accounting for circularity when it exists. Therefore, these methods are subject to serious, often unknown and unrecognized errors in stated probabilities associated with Type 1 error rates, loss of statistical power, or both. Statistical methods are presented that help circumvent these problems.

20. DISTRIBUTION / AVAILABILITY OF ABSTRACT
☒ UNCLASSIFIED/UNLIMITED ☐ SAME AS RPT. ☐ DTIC USERS

21. ABSTRACT SECURITY CLASSIFICATION
Unclassified

22a. NAME OF RESPONSIBLE INDIVIDUAL
Donald F. Haggard

22b. TELEPHONE *(Include Area Code)*
(502) 624-3450

22c. OFFICE SYMBOL
PERI-IK

DD Form 1473, JUN 86 *Previous editions are obsolete.*

SECURITY CLASSIFICATION OF THIS PAGE
UNCLASSIFIED

ARI Special Report 16

16. SUPPLEMENTARY NOTATION (Continued)

Triangle Park, NC 27709-2211. Donald F. Haggard, Contracting Officer's Representative.

Special Report 16

Circular Statistical Methods: Applications in Spatial and Temporal Performance Analysis

Robert P. Mahan
The University of Georgia

Field Unit at Fort Knox, Kentucky
Donald F. Haggard, Chief

Training Research Laboratory
Jack H. Hiller, Director

U.S. Army Research Institute for the Behavioral and Social Sciences
5001 Eisenhower Avenue, Alexandria, Virginia 22333-5600

Office, Deputy Chief of Staff for Personnel
Department of the Army

April 1991

FOREWORD

The U.S. Army Research Institute for the Behavioral and Social Sciences (ARI) Field Unit at Fort Knox is responsible for conducting research in Armor training and simulation and human performance in Armor weapon systems. Research on automated performance measurement systems has gained increasing importance as Armor training has come to rely more heavily on computer-based simulations. This research surveyed special statistical methods designed to analyze quantitative measures scaled as circular functions. The report illustrates the application of circular statistics to typical examples of performance data (i.e., orientation, navigation, time-of-day measures). These examples show that analysis of circular measures by standard linear statistical procedures can produce inaccurate or misleading information. Methods that are appropriate for circular measures are required in any automated performance measurement system.

This research was performed in the Summer Faculty Research and Engineering Program under a U.S. Army Research Office contract with the Battelle Memorial Institute. The research was sponsored by a Letter of Agreement (LOA) between ARI, the U.S. Army Armor Center and Fort Knox, the U.S. Army Materiel Command, and the U.S. Army Training and Doctrine Command effective 16 January 1989. The LOA, "Effects of Simulators and Other Training Resources on Training Readiness," identified needs for research on training methods using networked simulators. The research was related to ARI task 3204, Training Requirements for Combined Arms Simulators, part of the ARI Exploratory Development (6.2) Program Area 3, Training for Combat Effectiveness, in the Simulators and Training Devices product line.

The results of the survey and application of circular statistical procedures demonstrate methods useful in automated collective performance measurement systems for networked training simulators. These methods will contribute to research and development with a prototype Unit Performance Analysis System (UPAS) to collect and analyze spatial and temporal performance measures from the Simulation Networking (SIMNET) testbed system. The UPAS, working with SIMNET data, will be used to examine performance measurement requirements for future networked simulator systems. The report also should be of general methodological interest to agencies that conduct tests and evaluations or other research employing circular measures of spatial or temporal performance.

EDGAR M. JOHNSON
Technical Director

ACKNOWLEDGMENT

I would like to thank Dr. David Bessemer, U.S. Army Research Institute for the Behavioral and Social Sciences, Fort Knox, Kentucky, for his expert comments and guidance on earlier drafts of this manuscript.

CIRCULAR STATISTICAL METHODS: APPLICATIONS IN SPATIAL AND TEMPORAL PERFORMANCE ANALYSIS

EXECUTIVE SUMMARY

Requirement:

As part of an effort to develop measures for assessing leader and unit performance during tactical exercises conducted in networked simulators, this report surveys a group of procedures that can be used to evaluate spatial and temporal performance. The report focuses on certain aspects of tactical operations that involve directional angles, or that are periodic in time, and therefore are appropriately analyzed by circular statistical methods.

Procedure:

A literature review was conducted to find methods appropriate for the analysis of spatial and temporal data. Once several procedures were identified having potential relevancy to simulator data analysis, specific examples were then developed using these procedures to guide the analysis of simulation data based on circular scales. Examples in navigation performance, call-for-fire, and other directional events were evaluated within the context of simulation network battlefield training and development.

Findings:

The results in evaluating several circular statistical procedures indicate that there are a variety of possible problems that can emerge when using linear-based analyses on circular variables. For example, the usual methods for describing central tendency and variance are not appropriate for circular data. Another limitation in linear analysis concerns the issue of 'response bias' in directional measures on a plane. Traditional linear analyses may be insensitive to directional response biases, thus, potentially limiting the information generated from simulation network exercises useful in developing improved tactical systems and operational doctrine. Finally, various dimensions of statistical inference are altered when applying traditional nonparametric and parametric methods to circular data, and as a result, can produce misleading information on Type 1 error rates, statistical power, or both.

Utilization of Findings:

The statistical methods described here will be used by the U.S. Army Research Institute Field Unit at Fort Knox to develop automated performance measurement systems for spatial behavior in networked simulators. More generally, these methods should be used to supplant standard statistical methods whenever researchers obtain human or system performance data having circular scales.

CIRCULAR STATISTICAL METHODS: APPLICATIONS IN SPATIAL AND TEMPORAL PERFORMANCE ANALYSIS

CONTENTS

Page

LIST OF TABLES

Table

LIST OF FIGURES

Figure

CIRCULAR STATISTICAL METHODS: APPLICATIONS IN SPATIAL AND TEMPORAL PERFORMANCE ANALYSIS

Introduction

The majority of statistical techniques used in the analysis of human performance and training data are linear. The assumptions necessary in using linear statistical methods are often easy to specify and provide tractable mathematical solutions for modeling a wide range of events. However, it is indeed clear that many problems encountered in military training scenarios do not lend themselves to strict linear representation.

One unique system of measures that frequently cannot be modeled in a linear manner are data produced from circular scales. These variables are distinctive in the sense that data points are distributed on a circle instead of the traditional configuration of points on the real number line. Circular scales produce cyclic or periodic data that complicate traditional analytic procedures. The complexities found in evaluating circular data are largely a manifestation of the special interval level status the circular scale represents. Circular scales do not have a true zero point. In addition, the fact that they are circular means that any designation of high or low or more or less is purely arbitrary. For example, observations on a plane surface and rhythmic temporal phenomena can be viewed as being circular in nature, and thus appropriately analyzed by circular statistical methods.

Although biomathematics has long appreciated the idea of circular distributions in animal behavior studies on homing, migration, escape, and exploratory behavior to name a few, application in the human sciences remains minimal. Work in the general area of spatial and temporal performance, such as navigation, work-system design, biological rhythms, and sleep make issues on circular data analysis important to consider.

This paper attempts to survey some of the more basic ideas associated with the unique analytic problems that arise from directional measures on a plane. Although, Batschelet (1965, 1972, 1981) has pioneered many of the principles in statistical circular methods, some of his work is no longer in print and thus not readily available. Instead, one must rely on works in a variety of diverse areas including, biostatistics, animal behavior, factor analysis and circadian physiology to find examples on circular problems. The goal of the paper is to highlight procedures that may facilitate characterizing dimensions of circular data inaccessible to linear-based analyses. Several examples are presented that illustrate the usefulness of circular methods and similarly, help alert the researcher to logical problems along with the potential for information loss that may be encountered when analyzing periodic data using linear methods.

The context in which some of the examples are developed is taken from research efforts in simulator network training (SIMNET-T) and development (SIMNET-D). Briefly, SIMNET-T is a networked, distributed processing simulator AirLand battlefield developed to complement combined arms field training exercises in AirLand battlefield conditions. SIMNET-D is a reconfigurable SIMNET that provides a test-bed for prototyping futuristic weapons systems and operational doctrine. These simulations allow many players to engage in interactive, real-time battles against other human players or computer generated opposing forces at remote locations in the U.S. and Europe. Data from these battles can be collected and made available for analysis via the Unit Performance Analysis System Software (White, McMeel & Gross, 1990).

Mathematical Convention

Two different conventions are used with directional measures. The angle α is typically used as a measure of azimuth where 0 represents true north and rotation around the circle is in a clockwise direction. This convention is used in navigation.

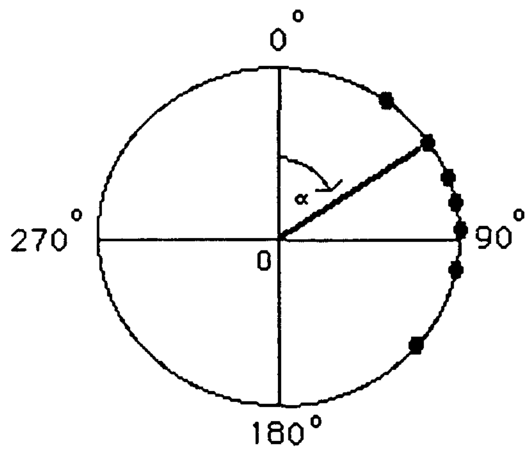

Figure 1. Directional angle computation showing clockwise rotation.

However, the mathematical convention for angular measures in statistics and computer computational algorithms typically use the polar angle, ϕ, which is taken from the positive X axis (pointing East) in a counterclockwise direction. This computational convention will be used in this paper unless otherwise indicated.

For both types of angles, rotation opposite to the conventional direction results in negative angles. Negative angles are subtracted from 360° to convert them to the

corresponding positive angles, unless it is important to retain knowledge of the direction of rotation.

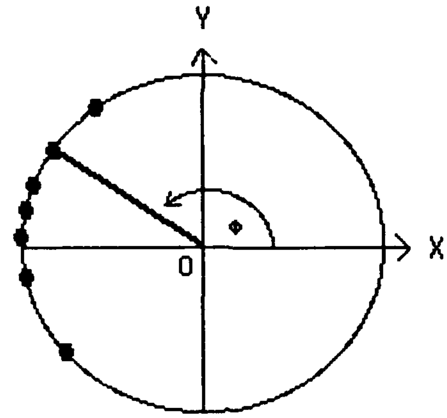

Figure 2. Mathematical angle computation showing counterclockwise rotation.

Temporal Measures

One particularly important circular scale is time-of-day. The time-of-day scale can be partitioned into 24 hours each representing equal intervals of time. Angular measures are *typically taken from midnight* or 12:00 AM. In equating 24 hours to 360°, each hour represents 15 of angle from zero at midnight (i.e., 360°/24). Any unit of time similarly can be equated to a proportional part of 360°.

We can convert time measures in any unit to angular direction (in degrees) by the following equation:

$$\alpha = \frac{(360)\,t}{\mu} \qquad (1.0)$$

where α is angular degrees, t is a quantity of time and μ is the number of equal interval time units (i.e., 24 for hour, 1440 for minute, etc.) representing one rotation around the circle.

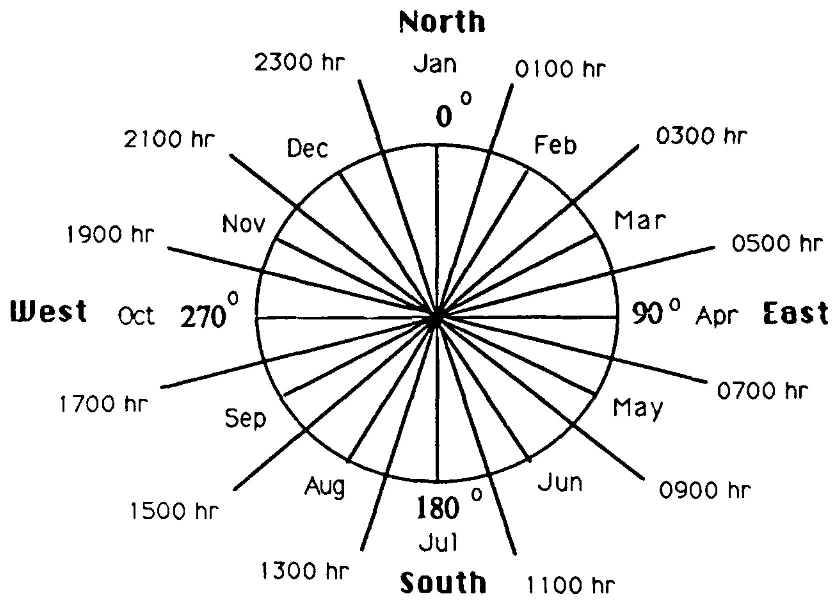

Figure 3. Three common circular measurement systems: Navigation direction, Time-of-day and Month-of-year.

Any circular temporal measure can be translated to angles using this method (i.e., day-of-week, day-of-year etc.). For example, 7 days would partition the circular scale of the week. Converting days to angular degrees would simply mean using 7 as the constant value, μ, in the above equation, with t measured in weeks. Finally, yearly measures would mean altering the μ value to reflect the appropriate unit of measure (i.e., 365 for days, 12 for months etc.).

When time is measured in negative and positive values for events before and after a zero point, the signs of negative angles should be retained to preserve information on the order of events. Then to make these angles positive for further computations, add an integral multiple of 360°. This corresponds to the addition of a constant to the time values, shifting the zero by an integral number of time cycles.

Graphical Representation

A circular data distribution can be displayed as a scatterplot of data points on the circumference of the circle. This method of presentation provides information on salient characteristics of the data, such as central tendency, dispersion (or concentration), and the number of modes that appear in the distribution. For example, in Figure 2, the data are

rather concentrated near 90° with no obvious modes.

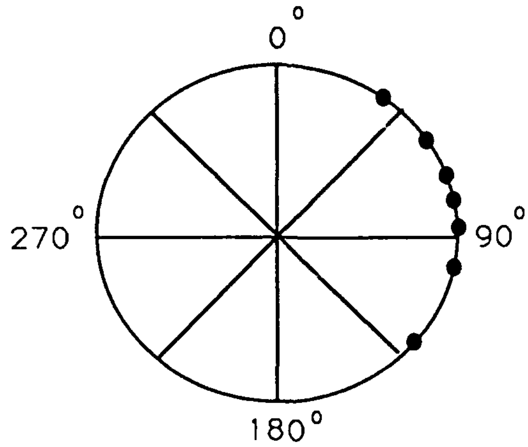

Figure 4. Scatterplot displaying circular data distribution.

Large amounts of data are best displayed in the form of a histogram where concentric circles characterize the frequency attributes of the data.

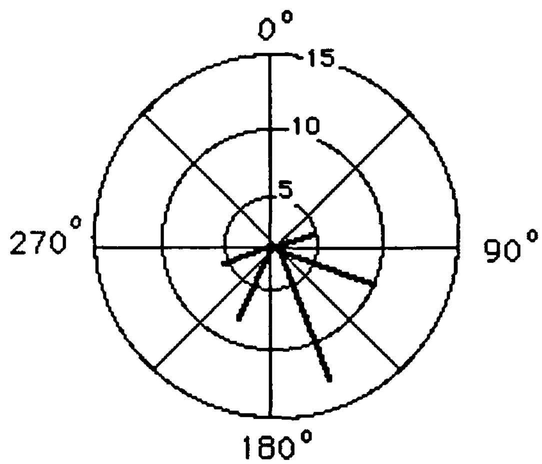

Figure 5. Frequency histogram of a circular distribution, using a grouping interval of 45°.

In this case, the length and area of the bars represents the frequency of observed measures. Similarly, Figure 6 displays a circular bimodal distribution.

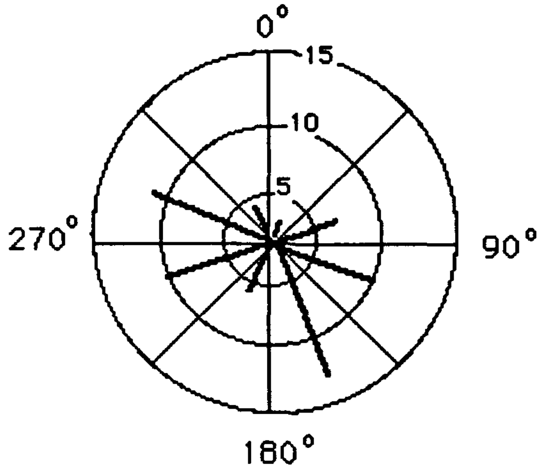

Figure 6. Frequency histogram showing bimodal circular distribution, using a grouping interval of 45°

Circular Functions

It is frequently necessary to change coordinate systems when analyzing circular data. The Cartesian system fixes a point on a plane by referring to the x and y axes of a rectangular representation of the event space. Rectangular coordinates read from the x and y axes give the unique location of the data point in the space. In Figure 7, point 1 is located on the unit circle by the rectangular coordinates x = 0.82 and y = 0.57, point 2 by x = -0.64 and y = 0.77, and so on. These points satisfy the circular equasion: $x^2 + y^2 = 1$

In many cases polar coordinates are more useful. The polar coordinate system requires the specification of the angle, ϕ, with respect to a zero or starting point and a straight line distance, r, from a second reference point located in the center of the circle representing the event space. The pair of numbers, ϕ and r, is known as polar coordinates and provide a method of uniquely defining the location of a data point in the circle. Take, for example, the four points in figure 7. If we specify the angle in respect to a northerly starting direction, and in this case travel <u>clockwise</u> around the circle, our polar values will be in azimuth (i.e. α and r). Thus, point 1 is given by $\alpha = 35°$, and r = 1.0, point 2 by α = 130°, r = 1.0 and so on.

More generally, polar coordinates can be used to specify points anywhere on a plane, not just on a unit circle. Instead of the rectangular coordinates (x_i, y_i), the polar

coordinates (ϕ_i, r_i) define the location of each point in terms of its direction and distance from the origin. Polar coordinates have the advantage of clearly separating directional and distance information in data analysis, while rectangular coordinates confound these aspects of spatial location. Rectangular coordinates, on the other hand, relate location to the frame of reference provided by the coordinate axes, preserving spatial information in terms of the orthogonal dimensions of the event space. The best choice of coordinate system for a particular problem depends on the questions to be investigated. Often both systems are useful for different purposes.

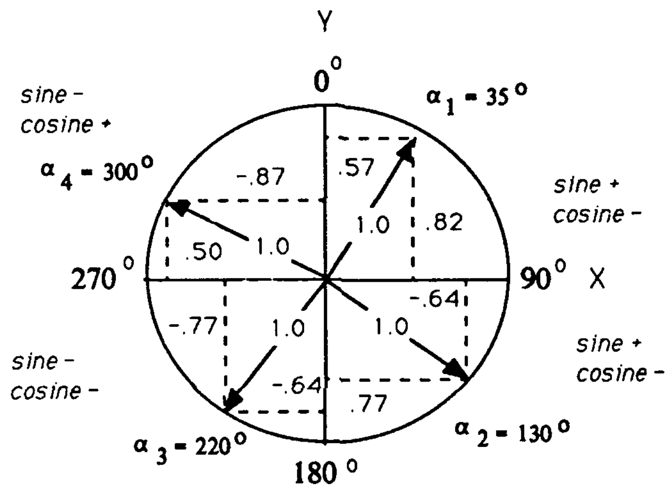

Figure 7. The unit circle displaying four data points with their rectangular (X and Y) and polar (α, and r) coordinates.

<u>Trigonometric functions.</u> The cosine of a particular angle is defined as the ratio of X and r for a circular value:

$$\cos \phi = \frac{X}{r}, \qquad (2.0)$$

whereas the sine of the angle is the ratio of the Y and the r for the circular value:

$$\sin \phi = \frac{Y}{r}. \qquad (3.0)$$

The sine of azimuth angle α on the unit circle is $\sin 35^\circ = 0.57/1.0 = 0.57$, and the cosine is $\cos 35^\circ = 0.82/1.0 = 0.82$.

Sines and cosines are the primary trigonometric functions used in circular statistical procedures. However, the tangent, and its cotangent function are used in several statistical tests one encounters in the literature:

$$\tan \phi = \frac{Y}{X} = \frac{\cos \phi}{\sin \phi} \qquad (4.0)$$

$$\cot \phi = \frac{X}{Y} = \frac{\sin \phi}{\cos \phi} \qquad (5.0)$$

It is important to recall that when analyzing circular data we restrict angular values to the interval length 0° - 360°. Since one rotation around the circle will contain the total set of angles, additional rotations can be mapped over to those angles defining this interval. For example, both 600° (1.66 revolutions) and -840° (negative 2.33 revolutions) determine the same point on the circle as the angle 240°. To find the angle in the 0° - 360° interval that corresponds to an angular value outside that interval, the value must be reduced modulo 360°. The angular value can be expressed as the sum of two parts: and integral multiple of 360°, $I(360^\circ)$, plus a remainder b after division by 360°. The remainder b is the desired angle within the interval. The angle $600^\circ = 1(360^\circ) + 240^\circ$, and thus reduces to 240°. Similarly, the angle $-840^\circ = -2(360^\circ) + -120$ and reduces to -120, which corresponds to the positive angle 240°. Both 600° and -840° are said to be congruent to the angle 240° (modulo 360°).

Azimuths and polar angles obey the relation $\alpha + \phi = 90^\circ$ (modulo 360°), since they sum to 90° in the first quadrant, or sum to 450° in other quadrants. Therefore, azimuths can be easily changed to polar angles by $\phi = (450^\circ - \alpha)$ (modulo 360°). With α and ϕ reversed, this equation will change polar angles to azimuths.

Descriptive Statistics

Circular Measures of Central Tendency

Linear descriptive statistics are limited in precisely characterizing the central tendency of circular distributions. For example, consider gun turret positions of a tank platoon in a staggered column formation given as 360°, 315°, 45° and 320° (see Figure 7). Determining the average turret azimuth by calculating the arithmetic mean would yield (360° + 315° + 320°)/4 = 260° which indicates a mean westerly direction. In contrast, the data points indicate that an average turret azimuth should produce a value reflecting a northerly direction. Thus, the usual arithmetic mean is not an appropriate descriptive measure for circular data.

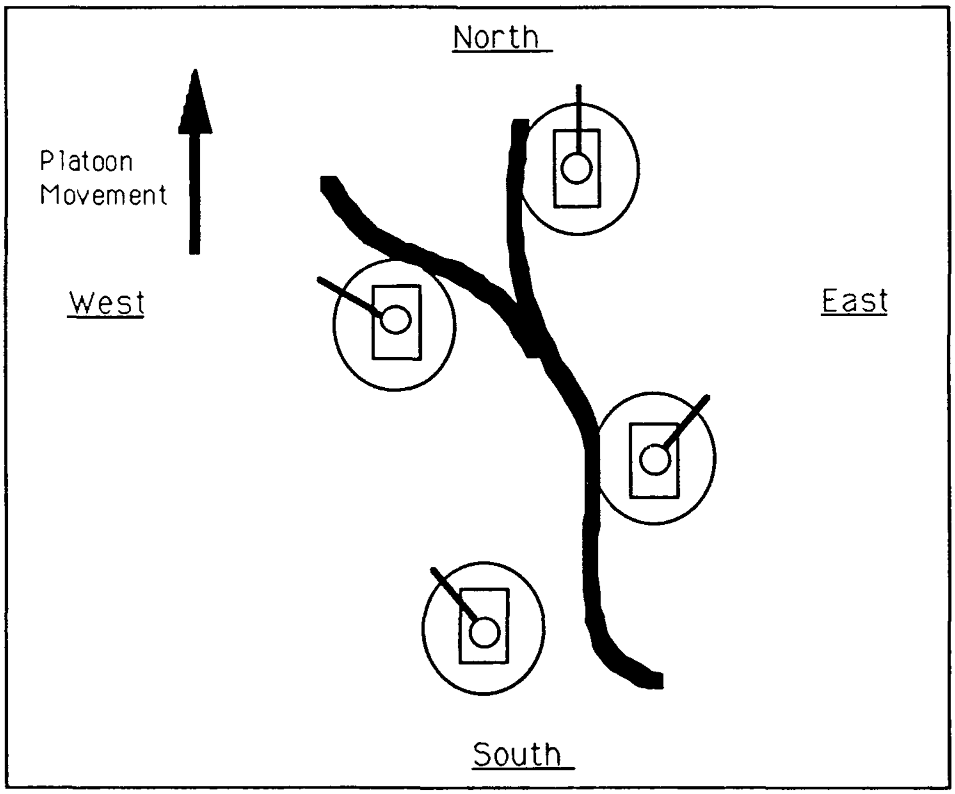

Figure 8. Staggered column formation showing northerly turret azimuth.

Circular mean. If we have a number of angular measurements on the circle, then the mean of those measures should offer an estimate of the true population mean parameter μ (in the classic probability sense). However, in order to compute the mean angle, the rectangular coordinates of the data points must be computed:

$$M_x = \overline{X} - \frac{\sum_{i=1}^{n} \cos \phi_i}{n} \qquad (6.0)$$

and

$$M_y = \overline{Y} - \frac{\sum_{i=1}^{n} \sin \phi_i}{n} \qquad (7.0)$$

These values allow for the computation of the length of the mean vector, r,:

$$r = [(\overline{X})^2 + (\overline{Y})^2]^{1/2} . \qquad (8.0)$$

The average angle, ϕ, is determined by computing sine and cosine values using X, Y and r:

$$\cos \overline{\phi} = \frac{\overline{X}}{r} \qquad (9.0)$$

$$\sin \overline{\phi} = \frac{\overline{Y}}{r} \qquad (10.0)$$

$\overline{\phi}$ can then be obtained from the inverse tangent function:

$$\overline{\phi}_i = \left| \begin{array}{ll} \text{arc tangent } (y/x) & \text{if } x > 0 \\ 180^{0} + \text{arc tangent } (y/x) & \text{if } x < 0 \\ 90^{0}, & \text{if } x = 0 \text{ and } y > 0 \\ 270,^{0} & \text{if } x = 0 \text{ and } y < 0 \\ \text{indeterminate,} & \text{if } x = 0 \text{ and } y = 0 \end{array} \right| \tag{11.0}$$

When $\overline{X} = 0$ and $\overline{Y} = 0$ then the length of the mean vector, r, is equal to 0, division by zero is undefined, and the mean direction is indeterminate.

As an example, consider the following turret azimuth data taken from one tank during a simulation exercise. In Table 1, turret azimuths are first converted to polar angles, and then these angles are used to compute the mean.

Table 1.
Tank turret angle data

Turret Angle (α)	ϕ	sine ϕ_i	cosine ϕ_i
59	31	0.515	0.857
37	53	0.797	0.602
37	53	0.797	0.602
355	95	0.992	-0.087
206	244	-0.898	-0.438
185	265	-0.996	-0.087
165	285	-0.965	0.258
164	286	-0.961	0.275

$\Sigma \text{sine } \phi_i = -0.719, \quad \Sigma \text{cosine } \phi_i = 1.982$

$$n = 8, \ \overline{Y} = -0.089, \ \overline{X} = 0.247$$

$$r = [(-0.089)^2 + (0.247)^2]^{1/2} = 0.296$$

$$\sin \bar{\phi} = \frac{-0.089}{0.296} = -0.300$$

$$\cos \bar{\phi} = \frac{0.247}{0.296} = 0.834$$

The average angle $\bar{\phi}$ corresponding to this sine and cosine is equal to 340.19°.

Median Angle

Computing the median angle of unimodal circular data is similar to that for linear data. In this case, we divide the the circular sample by a diameter that produces two equal sized groups of data on both sides of the diameter. When there are an odd number of data points, then the median is fixed on the datum such that (n - 1)/2 of the points lie on one side and the other (n - 1)/2 points on the other side of this unique location. If there is an even number of data points, then the median is located halfway between two points in such a way that (n/2) points fall on one side and the other (n/2) points fall on the other side of this location.

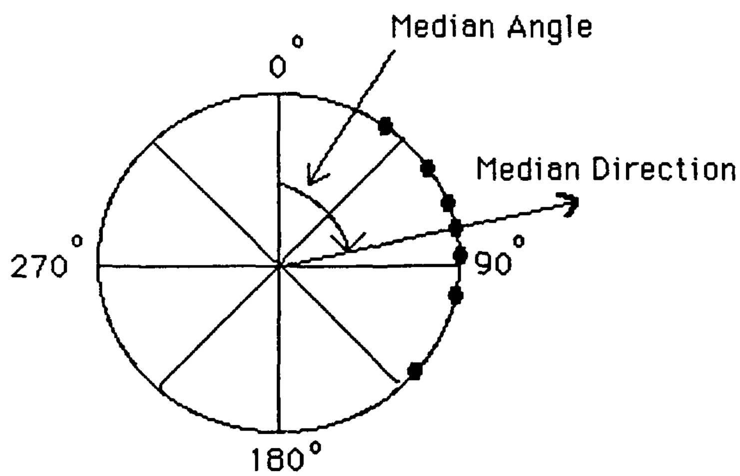

Figure 9. Circular distribution displaying median direction and angle.

Bimodal Samples

In some cases a mixture of unimodal distributions appear in a single sample. The most tractable, from a statistical point of view, are those referred to as axial distributions. In an axial distribution, two groups of data lie diametrically opposite each other on the diameter of the circle. Thus, the assumption is made that the probability density from which the data are sampled has central symmetry, with modes at μ and $\mu + 180^\circ$. We are indifferent to direction from the origin along the axis, and are only interested in the angle of the axis. We want to estimate one parameter μ rather than two parameters for the separate modes. It is therefore appropriate to treat angular measures as axial whenever orientation without regard to direction is at issue.

Batschelet (1965, 1981) discusses a procedure for analyzing axial data that requires doubling angular measures. One of the animal behavior studies cited by Batschelet (1981) notes that pigeons try to avoid large bodies of water. When pigeons are released from the center of a large body of water their flight route tends to be at right angles to the long axis of the lake, and thus the nearest shore. In this case, two sets of data lying 180° away from each other are recorded .

In analyzing axial data, one considers a nondirected rotating line segment whose origin lies at zero direction. A rotating line segment is restricted to angles in the interval of 0° to 180°, as opposed to the directed line that has the full 0° to 360° interval. As a result, axial data can be reduced by a multiple of 180°, or *modulo* 180°. One treats a full rotation around the circle as falling in the interval of 0° to 180°. Thus, by doubling each angle and reducing *modulo* 360°, one generates a unimodal circular sample (Taylor & Auburn, 1978).

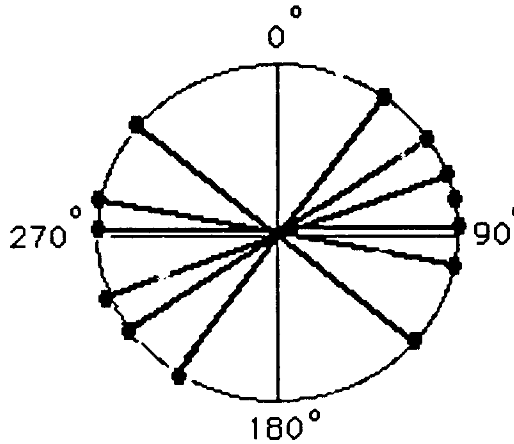

Figure 10. Bimodal (axial) circular distribution.

Properties of Mean Vector

Because we are working with data points falling on a circle, the mean vector locates the center of mass of the distribution. The mean vector is a directed line from the origin to the point defined by the Cartesian coordinates $\begin{bmatrix} \bar{X} \\ \bar{Y} \end{bmatrix}$, or the polar coordinates $(r, \bar{\phi})$. This notion can be illustrated by applying vector algebra in order to determine center of mass. For example, consider four data points falling on a unit circle. Each of these points can be located on the circle by a unit vector, υ, (i.e. directed line of length 1.0). Using a physical analogy, each point is associated with a value representing its mass, M, and location in space. The mean vector for these data points would be found by combing vector information (i.e., rectangular or polar coordinates) with each point's mass value:

$$\overline{M} = \frac{\sum_{i=1}^{n} M_i \times \upsilon_i}{\sum_{i=1}^{n} M_i} \qquad (12.0)$$

where M is the mass of each data point and υ is a column vector fixing the point's location in space. Equation (12.0) computes the weighted mean vector accounting for differences in

mass associated with the individual data points. If one assumes equal mass among data points, which would be a typical framework in which to evaluate a variety of circular data, then Equation (12.0) simplifies to:

$$\overline{M} = \frac{\sum \upsilon_i}{n} \qquad (13.0)$$

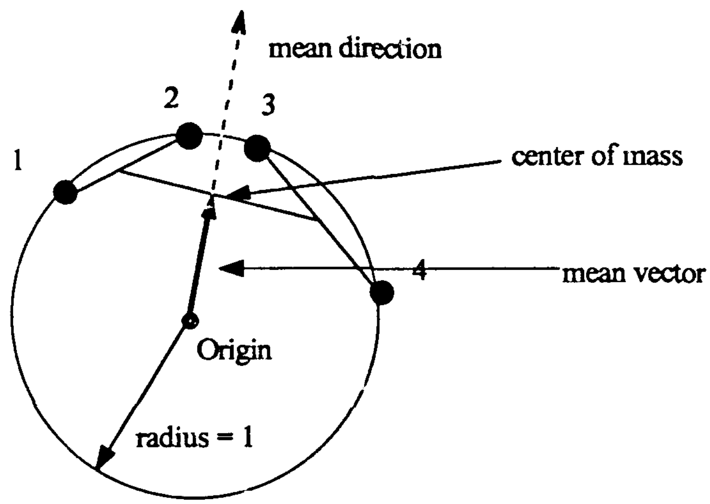

Figure 11. Four points of equal mass are displayed where the mean vector meets a line segment connecting the midpoints on two additional segments among the points.

Angular Variance

It is often useful to have a measure of dispersion around the mean vector in a sample of circular data. The notion of angular variance is similar to that of i's linear counterpart, namely, the quantity that defines the spread of scores around the circle's circumference. The length of the mean vector, r, plays an important role in the variance estimate. Recall from the vector algebra example, that as the data becomes more dispersed around the circle the value of r tends toward zero. When calculating the mean vector of a distribution whose data points are equally dispersed around the circle, the length of r will be zero. In contrast, when computing the mean vector of sample data whose points all fall on the same location, yields an $r = 1.0$. In this case, as r decreases from 1.0 to 0, the variance in the distribution increases. Thus, 1-r can be considered a measure of dispersion (Batschelet, 1972).

Batschelet (1981) describes several statistical similarities between linear and circular measures, which he indicates aids in the development of circular methods.

Linear statistics	Circular statistics
$(X_i - \bar{X})$	$\sin(\phi_i - \bar{\phi})$
$\sum (X_i - \bar{X}) = 0$	$\sin(\sum \phi_i - \bar{\phi}) = 0$
$(X_i - \bar{X})^2$	$2[1 - \cos(\phi_i - \bar{\phi})]$
$\frac{1}{n} \sum (X_i - \bar{X})^2 = s^2$	$\frac{1}{n} \sum 2[1 - \cos(\phi_i - \bar{\phi})] = 2(1-r)$

The analogy between linear variance and angular variance shows that the actual (radial) estimate is $2(1 - r)$. Taking the positive square root of this value (i.e.,$[2(1 - r)]^{1/2}$) results in the angular deviation. In this case, angular deviation is expressed in radians (i.e., $2\pi r = 360^{\circ}$). A simple transformation can convert the deviation into degrees (Batschelet, 1965):

$$s_{degrees} = \frac{180^{\circ}}{\pi} [2(1 - r)]^{1/2} \qquad (14.0)$$

In addition, there are other methods by which to compute the circular standard deviation (see Mardia, 1972). Equasion (14.0) yields a measure that ranges from 0° to 81.03°, or 0 to 1.41 radians.

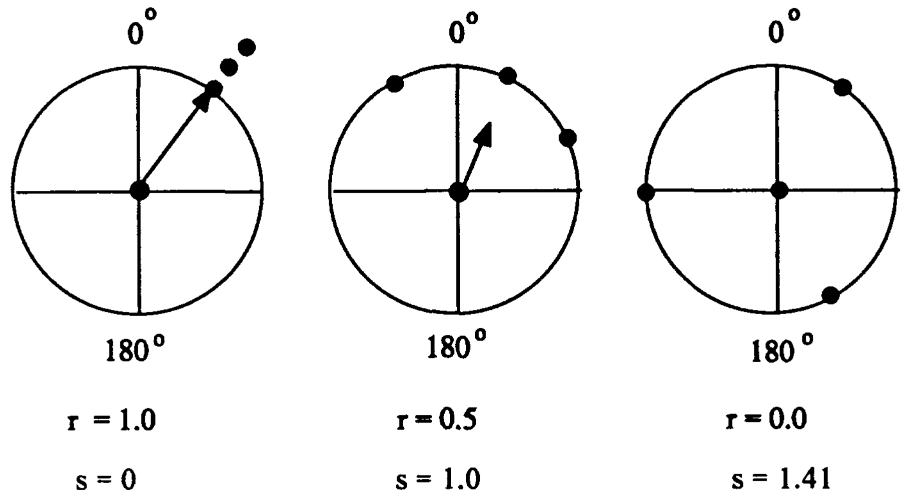

Figure 12. Measure of concentration showing the effect of dispersion on the statistic r, and angular deviation s (measured in radians).

Index of straightness. Although, an index of straightness may have limited utility in tactical exercises where terrain features (among other things) drive route objectives, the statistic r may be useful in providing information on deviations from planned route. This index may be very useful when modeling Naval exercises at sea where terrain is not a modeling constraint.

An index of straightness can be used to get approximate standard deviation values (s) for each segment of a route composed of straight lines when the cumulative distance traveled has been measured, but where headings and positions have not been measured continuously or at small intervals of time. For each route segment, the planned mean heading is obtained from the the segment start and end points. The actual mean heading must be computed from some defined points that correspond to the start and end of the segment, like actual positions A and B, measured when checkpoints at the start and end segment were reported. If the straight-line distance from A to B was C, and the distance traveled was D , then r *approximates* C/D, and s is obtained from r by Equation 14.0. The planned and actual mean headings, and r or s values then can be used for further analyses. However, these r or s values certainly will have sampling properties that differ from r or s values computed from a set of vectors.

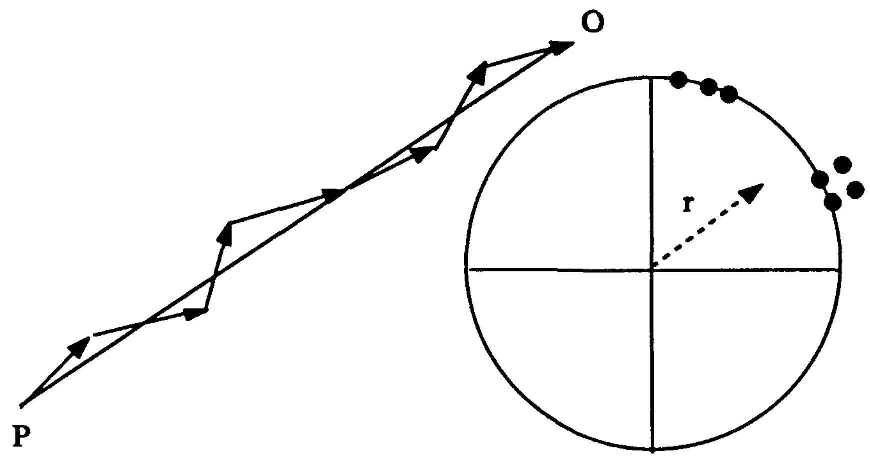

Figure 13. Mapping a linear vector series to a circle showing the effects of deviations from a straight heading on the statistic r. Here, $\overrightarrow{PO}$ gives a mean directional angle, and r approximates the value for 'length divided by distance', providing a measure of concentration.

Grouped data. The mean angle of grouped frequency data is often necessary to compute. In this case, taking the midpoint angle of a group interval and multiplying it by the frequency of a measure falling into the interval is similar to the approach taken with linear statistics:

$$\overline{X} = \frac{\sum_{i=1}^{n} fi \cos \phi_i}{n} \qquad (15.0)$$

and

$$\overline{Y} = \frac{\sum_{i=1}^{n} fi \sin \phi_i}{n} \qquad (16.0)$$

Note that $n = \sum_{i=1}^{n} fi$.

However, the mean vector length, r, underestimates the true population ρ, thus, a correction for this bias has been proposed by Batschelet (1965). Here, the value of r

$$c = \frac{\frac{d\,\pi}{360^{\circ}}}{\sin\left(\frac{d\,\pi}{360^{\circ}}\right)} \qquad (17.0)$$

where d is the arc length of the class interval in degrees, c is the correction term and r_c is the corrected r,

$$r_c = cr\,. \qquad (18.0)$$

The corrected r is then used in Equation 14.0 to obtain the corrected angular deviation s_c. However, since r appears in both the average cosine and sine (equations 9.0 and 10.0), neither the tangent (equation 4.0) nor the resulting mean angle (equasion 11.0) are affected by the correction.

Detecting bias. One particular application of circular statistics may be found in assessing response biases on circular performance measures. Frequently, errors of judgment in navigation, orientation or other spatial performance tasks are reported as absolute or relative deviations from some true score (see Du Bois & Smith, 1989; Fisicaro, 1989). Rendering experimental conclusions based on linear deviations of circular measures from a true score, where deviations fall on both sides of azimuth, means that directional information is lost. For example, Fisicaro (1989) conducted a study of the performance characteristics of four orientation indicators that might accompany the CITV (Commander's Independent Thermal Viewer). He compared several orientation displays designed to indicate the directions of the tank hull, turret/main gun and the tank commander's sight relative to the surrounding terrain. Operator performance measured either (a) judged orientation of any one of these three tank components compared to its true orientation, or (b) the absolute deviation of the difference in judged orientation between pairs of component displays from the actual 'true' differences. Judgments were made in terms of eight compass directions.

However, these absolute deviation values do not indicate whether errors are symmetric around the true value, or whether there is a systematic tendency (bias) to make errors in one direction. Furthermore, the bias, if any, may be different at different points of the compass. For example, there might be a "forward-looking" bias, so that differences between the tank hull and the TC's sight are underestimated for angle differences between $\pm$ 60°, but not for other differences. Thus, the use of average absolute deviations to

measure directional errors of judgment can discard important information on directional bias.

The notion of response bias in spatial tasks have been an important dimension in the development of aviation navigation and remote vehicle guidance displays. Frame of reference appears as a key element and has given rise to various frame of reference configurations (eg. inside-out and outside-in). In addition, detecting direction errors in judgments occurring in spatial performance tasks may be very important because of the "rectilinear normalization" bias (Wickens, 1984). This bias characterizes a tendency of human subjects to structure spatial information as though lying on a rectangular N-S-E-W grid, when it may not.

Inferential Statisitcs

Traditional linear-based methods of statistical inference do not take into account scale circularity when it exists, and therefore can be subject to unknown and unrecognized errors in specified probabilities of Type 1 error, loss of statistical power, or both. Statistical methods are avialable for circular data that minimize these problems, if their assuptions are met.

The von Mises Distribution

A frequently used theoretical distribution for fitting circular data is the von Mises distribution, which was defined by von Mises in 1918. It is best suited to modeling symmetric unimodal circular data distributions. A two parameter density function characterizes this distribution:

$$f(\phi) = \frac{1}{2\pi I_o(k)} \exp\left[k \operatorname{cosine}(\phi - \theta_1)\right] \qquad (19.0)$$

where k is a concentration parameter estimated by the length of the mean vector, r, and θ is the polar angle of the mean vector μ. Here, the quantity, I_0, found in the denominator of the equasion is a Bessle function parameter. Bessel functions have certain important properties for circular methods (see Batschelet: 295 - 299, 1981). Batschelet (1972, 1981) notes that the density function takes on a maximum value at $\phi = \theta_1$, thus, θ_1 is the mode of the distribution. Because the distribution is symmetric around the mode, θ_1 is the mean angle. When $k = 0$ this distribution is equivalent to the uniform distribution. Tabled values for the von Mises distribution are found in Batschelet (1972, 1981).

Uniform Circular Distribution

If directions in a plane can occur with equal probability, then the distribution of data points around the circumference of the circle will tend to be uniform. Thus, the density of data falling around the circle will be constant. The uniform circular distribution is a good model for many random circular stochastic processes. It provides the probability density used to test the null hypothesis of 'no preferred direction'.

Several authors have noted that the uniform circular distribution has unique properties. However, one of the more interesting is that it appears as the only circular distribution where, in random samples of a fixed size, the length of the mean vector and therefore the

angular variance is statistically independent from the mean angle (Batschelet, 1972; Bingham, 1978).

Confidence Interval for Mean Angle

Characterizing the confidence with which a particular estimate closes on the population parameter is important. Tabled values for 95% and 99% confidence limits can be found in Batschelet *(1965, 1972, 1981) and Zar (1974).* In these tables, a given vector length, r, and sample size, n, provides the quantity, *d*, which reflects the confidence interval for the parameter μ_a:

$$\bar{\alpha} \pm d \qquad (20.0)$$

We assume that the the sample is drawn from a population with a von Mises distribution, and the sample is unimodal and symmetric about the mode. Here, the lower confidence limit is equal to $L_L = \bar{\alpha} - d$, and the upper confidence level equal to $L_u = \bar{\alpha} + d$.

The Grand Mean

It is often necessary to compute the grand mean from a sample of group means. However, in computing the grand mean of several means, it is not accurate to consider each of the group means as an angular measure, and divide the sum of mean angles by n. This procedure would assume the unlikely situation that the mean vector length of each group mean vector was equal to 1.0 and that the angular deviation for each group was equal to zero. Instead, we fix the location of the group means via cartesian or polar coordinates,

and then sum across the respective coordinates:

$$\overline{X}_{gm} = \frac{\sum_{j=1}^{k} \overline{X}_j}{k} \qquad (21.0)$$

$$\overline{Y}_{gm} = \frac{\sum_{j=1}^{k} \overline{Y}_j}{k} \qquad (22.0)$$

where X and Y are the rectangular coordinates for the sample means, and k is the number of samples. The above equations compute an unweighted mean, thus sample sizes must be reasonably equivalent.

Similarly, if we have the polar coordinates for each of the group means, the grand mean coordinates are computed as follows:

$$\overline{X}_{gm} = \frac{\sum_{j=1}^{k} r_j \cos \overline{\Phi}_j}{k} \qquad (23.0)$$

and

$$\overline{Y}_{gm} = \frac{\sum_{j=1}^{k} r_j \sin \overline{\Phi}_j}{k} \qquad (24.0)$$

The equations for computing the grand mean vector length are essentially identical to those for computing the mean vector length described in Equations (9.0 and 10.0).

$$r_{gm} = [(\overline{X}_{gm})^2 + (\overline{Y}_{gm})^2]^{1/2} \quad (25.0)$$

and

$$\cos \overline{\phi}_{gm} = \frac{\overline{X}_{gm}}{r_{gm}} \quad (26.0)$$

$$\sin \overline{\phi}_{gm} = \frac{\overline{Y}_{gm}}{r_{gm}} \quad (27.0)$$

The grand mean angle would then be obtained from the inverse tangent as before (i.e., Equation 11.0).

As an example in computing the grand mean, consider the mean direction of travel for three simulated tank platoons engaged in a coordinated tactical mission to close on enemy positions. Computing the grand mean direction of travel at differing intervals of time during the simulation might offer information that can be used in secondary analyses as to how well the platoons were able to achieve planned route objectives.

Table 2.
Directional data of a tank company taken at times T_1 to T_5

Platoon Direction of Travel (degrees)

	Platoon 1			Platoon 2			Platoon 3		
	dir	sin	cos	dir	sin	cos	dir	sin	cos
T_1	270	-1.000	0.000	190	-0.174	-0.984	95	0.996	-0.087
T_2	245	-0.900	-0.423	180	0.000	-1.000	40	0.642	0.766
T_3	220	-0.643	-0.766	200	-0.342	-0.939	130	0.766	-0.643
T_4	260	-0.985	-0.174	200	-0.342	-0.939	170	0.174	-0.985
T_5	265	-0.996	-0.087	225	-0.707	-0.707	110	0.939	-0.342
Σ		-4.524	-1.450		-1.565	-4.569		3.517	-1.291

$\phi = 252^{\circ}$	$\phi = 199^{\circ}$	$\phi = 109^{\circ}$
$r = 0.950$	$r = 0.966$	$r = 0.749$
$\sin\phi = -0.90$	$\sin\phi = -0.325$	$\sin\phi = 0.940$
$\cos\phi = -0.290$	$\cos\phi = -0.914$	$\cos\phi = -0.258$

We can now compute the grand mean angle by substituting the intermediate terms from the directional data using Equations (25.0 - 29.0)

$$\overline{Y}_{gm} = \frac{(.950)\ -.90 + (.966)\ -.324 + (.749)\ .940}{3} = -.322$$

$$\overline{X}_{gm} = \frac{(.950)\ -.290 + (.966)\ -.914 + (.749)\ .258}{3} = -.155$$

$$r_{gm} = [(-.155)^2 + (-.322)^2]^{1/2} = .357$$

$$\cos\overline{\phi}_{gm} = \frac{-.155}{.357} = -.434$$

$$\sin\overline{\phi}_{gm} = \frac{-.322}{.357} = -.90$$

Therefore, $\overline{\Phi}_{gm}$ is equal to 244.25°.

When the sample sizes are unequal, equations (23.0 and 24.0) must be modified to weight each mean by $(n_i / \Sigma\, n_i)$.

Confidence Interval for Grand Mean

Similar to the confidence interval discussed for the mean angle of a sample of angles, one can ask questions concerning the precision of the grand mean estimation of the parameter, $\overline{\Phi}_{gm}$. In this case, one must first determine that the data are not uniformly distributed on the scale and that significant directionality exists. Batschelet (1981) provides a detailed examination of confidence limits which requires knowledge on the geometry of ellipses. Here we present a cookbook procedure for determining the limits (see Zar, 1974). The confidence limit obtained when calculated in this manner tends to be conservative (i.e., greater than the $1 - \alpha$ convention). Furthermore, it may not be symmetrical about the mean (Zar, 1974). This method assumes that the mean vectors have a bivariate normal distribution.

The following parameters are found when solving for a confidence ellipse (Batschelet, 1981: 129-158) and is reported in Zar, (1974):

$$A = \frac{k - 1}{\Sigma x^2} \qquad (30.0)$$

$$B = \frac{(k - 1) \ \Sigma xy}{\Sigma x^2 \ \Sigma y^2} \qquad (31.0)$$

$$C = \frac{k - 1}{\Sigma y^2} \qquad (32.0)$$

$$D = \frac{2(k - 1) \ \left[1 - \frac{(\Sigma xy)^2}{\Sigma x^2 \ \Sigma y^2}\right] F_{\alpha(1), 2, k-2}}{k(k - 2)} \qquad (33.0)$$

$$H = AC - B^2 \quad (34.0)$$

$$G = A\overline{X}^2 + 2B\overline{XY} + C\overline{Y}^2 D \quad (35.0)$$

$$U = H\overline{X}^2 - CD \quad (36.0)$$

$$V = (DGH)^{1/2} \quad (37.0)$$

$$W = H\overline{XY} + BD \quad (38.0)$$

$$b_1 = \frac{W + V}{U} \quad (39.0)$$

$$b_2 = \frac{W - V}{U} \quad (40.0)$$

The values b and b are analyzed individually, thus producing the bounds of the confidence interval for the grand mean as follows:

$$M = (1 + b_i^2)^{1/2} \quad (41.0)$$

$$\text{sine} = b_i / M \quad (42.0)$$

$$\text{cosine} = 1 / M \quad (43.0)$$

The confidence interval is the angle computed, or the angle <u>plus</u> 180°, (which ever is closest to the grand mean) as is the appropriate value given by the convention for angular measures on a circle:

$$\phi = \begin{cases} \text{arc tangent } (y/x) & \text{if } x > 0 \\ 180^{\circ} + \text{arc tangent } (y/x) & \text{if } x < 0 \end{cases} \qquad (44.0)$$

In this case, Zar (1974) notes that if the angle for the confidence limit is equal to (angle + 180°), and if the resulting angle is greater than 360°, one subtracts 360° from this angle.

Significance Tests

One of the more frequently needed assessments on experimental data is determining whether obtained sample data has been generated from a population distribution different than one generating a random distribution of measures. In circular data, the question may be whether a mean direction in the population data exits. This question is similar to one found in a linear system, except the null hypothesis characterizes a uniform distribution (i.e., equal density of points on the circle) that can yield a computable mean by chance alone. A simple nonparametric test for the competing hypotheses; a) H_0: the data comes from a uniformly distributed population, and b) H_1: the data comes from a non-uniform circular distribution, is the Rayleigh test.

The Rayleigh test essentially asks how large the statistic r must be in order to indicate a significant departure from uniformity. The "Rayleigh R" is computed as follows:

$$R = nr, \qquad (45.0)$$

where, n = to sample size and r is = to the mean vector length of the sample data. The Rayleigh z score is used to test the null hypothesis of uniformity:

$$z = \frac{R^2}{n}. \qquad (46.0)$$

Tabled values for this test are found in Batschelet (1965, 1972, 1981) and Zar (1974). In addition, one can compute the probability function using Durrand and Greenwood's (1958) formulation.

If the null hypothesis is rejected using the above test, we may assume that significant directionality exists in the data. In order to interpret the data in this fashion, we must assume that the distribution is unimodal. Similarly, in retaining the null and concluding the

data has been sampled from a uniform population distribution, caution must be used to guarantee the distribution is not multimodal. For example, an axial distribution, while displaying significant directionality, will have an $r = 0$.

Durrand and Greenwood (1958) provide a modification to the Rayleigh test which allows for the specification of an apriori expectation for a particular mean direction. In this case, the null hypothesis would indicate that the population of angles appear with equal density around the circle, or the concentration population parameter, r, equal to zero. In contrast, the alternate hypothesis proposes that r is not equal to zero and that a specific direction exists in the data.

In the example given above for computing the mean turret angle, consider that we had expected a mean angle of 270°. This alternate hypothesis would be tested by the V test. The V statistic is computed as follows:

$$V = R\cos(\overline{\phi} - \mu_o) \qquad (47.0)$$

where μ_o is the mean angle predicted. Tables for the V statistic are found in Batschelet (1972, 1981). The critical values for this statistic have been found to approximate a one-tailed normal deviate, Z. In this case, for $\alpha = .10, .05$, or $.01$, if $n \geq 5$ the one-tailed normal curve values 1.282, 1.645, or 2.236 can be used with a deviation less than 3% from the nominal α. For more extreme values of α, use Table 1 in Batschelet (1981, pp. 336).

$$\mu = V\left(\frac{2}{n}\right)^{1/2} \qquad (48.0)$$

Table 3.
Tank turret data

ϕ (degrees) Turret Angle	sine ϕ	cosine ϕ
31	0.515	0.357
53	0.797	0.602
53	0.797	0.602
95	0.992	-0.087
244	-0.898	-0.438
265	-0.996	-0.087
285	-0.965	0.258
286	-0.961	0.275

$$\Sigma \text{sine}\,\phi = -0.719, \quad \Sigma \text{cosine}\,\phi = 1.982$$
$$n = 7, \quad Y = -0.089, \quad X = 0.247$$

$$r = [(-0.089)^2 + (0.247)^2]^{1/2} = 0.296$$

$$\sin \bar{\phi} = \frac{-0.089}{0.296} = -0.300$$

$$\cos \bar{\phi} = \frac{0.247}{0.296} = 0.834$$

$$\bar{\phi} = 340.19°.$$

Rayleigh's R and the V test:

$$R = (8)(0.296) = 2.368$$

$$V = R \cos(340.19^{\circ} - 270^{\circ})$$

$$= 2.368 \cos(70.19^{\circ})$$

$$= 0.802$$

$$\mu = V \left(\frac{2}{n}\right)^{1/2}$$

$$= 0.802 \left(\frac{2}{8}\right)^{1/2}$$

$$= 0.401$$

The tabled value for V ($p < .05$, $n=8$) is equal to 1.649. Therefore, we retain the null hypothesis that no mean directionality exists in the data, or that the population distribution is uniform in nature.

One-Sample Mean Angle Test

The Rayleigh and V tests are appropriate for testing the hypothesis of a random distribution of data points on a circle. However, if one is interested in determining whether the population mean angle is equal to some specified value, then we use the one-sample mean test. The one-sample test can be considered an analog to the t-test in linear statistics. The procedure used for the one-sample test is simply the determination of whether or not our observed angle lies within the $1 - \alpha$ confidence interval for our specified population mean angle. The procedure is essentially the same for determining the confidence limits for the mean population angle described earlier. Values for the r statistic and the observed mean sample angle are used to identify the tabled value of *d* for the confidence interval. If the confidence interval doe not contain the hypothesized mean, then we reject the null hypothesis of the sample mean being equal to some specified population value:

H_o: The population mean has a mean of X, (i.e., $\mu_{\Phi} = X$)

v.s.

H_1: The population mean is not equal to X, (i.e., μ_{Φ} not equal to X).

Multiple-Sample Mean Angle Test

An extension of the one-sample mean angle test is found when one considers the null hypothesis that the mean angle from one sample is equal to the mean angle of a second or third sample and so on (i.e., $\mu_{\bar{\Phi}1} = \mu_{\bar{\Phi}2} = \ldots\ldots \mu_{\bar{\Phi}k}$). The Watson-Williams test (1956) uses a modification of the linear F-test replacing the standard group mean estimates with Rayleigh's R values generated from the sample data. The two sample parametric F test follows:

$$F = K \frac{(N-2)(R_1 + R_2 - R)}{N - R_1 - R_2} \qquad (49.0)$$

where N is the combined sample size, R_1 and R_2 are Rayleigh's R values computed independently for the samples, and R is Rayleigh's R value computed for the combined sample data. The K factor is a tabled value that corrects for the bias in the F-test and and can be found in Batschelet (1972,1981), Zar (1974). It can also be computed from equations developed by Mardia (1972). In using the K correction factor for the Watson-Williams test, one computes a weighted mean vector length:

$$\bar{r} = \frac{n_1 r_1 + n_2 r_2}{N} \qquad (50.0)$$

As an example, consider the gun elevation angles among four simulated tank platoons engaging enemy forces in a simulation network exercise. In this exercise, there were 16 tanks in a degraded-mode gunnery operation, each firing at the same target, but using two different methods of estimating target range to select the aiming point on a sight reticle. Since the range estimates were not recorded, gun elevation was used as a directly related substitute measure to determine if the average range was different for the two methods.

Table 4.
Gun elevation angles taken from four tank platoons

Group #1			Group #2		
Gun Angle (degrees)	sine	cosine	Gun Angle (degrees)	sine	cosine
37	.60181	.79863	22	.37460	.92718
19	.32556	.94551	20	.34202	.93969
20	.34202	.93969	16	.27563	.96126
67	.92050	.39073	44	.69465	.71934
13	.22495	.97437	39	.62932	.77714
122	.84804	-.52991	52	.78801	.61566
40	.64278	.76604	87	.99863	.05233
29	.48481	.87462	12	.20791	.97814

$n_1 = 8$ $\Sigma\text{sin} = 4.39$ $\Sigma\text{cosin} = 5.16$ $n_2 = 8$ $\Sigma\text{sin} = 4.31$ $\Sigma\text{cosin} = 5.97$

$\overline{Y} = .548$ $\overline{X} = .645$ $\overline{Y} = .538$ $\overline{X} = .746$

$r_1 = .846$ $r_2 = .919$

$\sin \Phi_1 = .6477$ $\text{cosin } \Phi_1 = .7624$ $\sin \Phi_2 = .5854$ $\text{cosin } \Phi_2 = .8117$

$\Phi_1 = 40^\circ$ $\Phi_2 = 36^\circ$

$R_1 = 6.768$ $R_2 = 7.352$

Combining the values from both groups yields:

$$\Sigma\text{sin} = 4.39 + 4.31 = 8.70$$

$$\Sigma\text{cosin} = 5.16 + 5.97 = 11.13$$

$$N = 8 + 8$$

$$Y = \frac{8.70}{16} = .5437$$

$$X = \frac{11.13}{16} = .6956$$

$$r = .8828$$

$$R = 14.124$$

$$\bar{r} = \frac{6.768 + 7.352}{16} = .8825$$

$$F = K\frac{(N - 2)(R_1 + R_2 - R)}{N - R_1 - R_2}$$

$$= 1.0823 \frac{(16 - 2)(6.768 + 7.352 - 14.124}{20 - 6.768 - 7.352}$$

$$= 1.0823 \frac{-.0560}{5.88} = -.0103$$

$$F, 0.05, 1, 14 = 4.60$$

Therefore, we do not reject the null hypothesis and conclude there is no statistical difference among gun elevation angles between the two groups.

The multi-group Watson-Williams test is a generalized form of the two group test configured for k groups:

$$F = K \frac{(N-k)\left(\sum_{j=1}^{k} R_j - R\right)}{(k-1)\left(N - \sum_{j=1}^{k} R_j\right)} \qquad (51.0)$$

where, k is the number of groups, R_j is Rayleigh's R for each group, and R is Rayleigh's R combined across groups. N is the total number of elements across groups. The tabled correction factor K is found using:

$$\bar{r} = \frac{\sum_{j=1}^{k} n_j r_j}{N}. \qquad (52.0)$$

Time Series Analysis of Periodic Data

Perhaps, one of the more difficult circular statistical procedures is encountered analyzing data where the basic unit of measurement is a temporal sequence of observations across time. In time series analysis, the somewhat restrictive assumption that sources of individual and measurement error variance can be modeled as a random process is not applicable. This is due to the fact that the analysis is based on repeated observations on the same subject.

The analysis of periodic time series data can take several forms depending on the types of variables included in the design. The different bivariate variable configurations can be a) circular dependent variable/linear independent variable, b) circular dependent variable/circular independent variable, or c) linear dependent variable/circular independent variable, where the independent variable is a fixed effect. However, one of the more statistically tractable variable combinations, and one routinely found in the literature, is the latter case; a linear dependent variable and circular independent variable. Frequently, a hypothesis to be tested using circular time series data is that the time series exhibits some sort of predictable regularity of period, and that the series begins repeating itself after a certain interval of time. This particular hypothesis is especially important in the field of chronobiology, which concerns itself with changes in many linear scaled biological/behavioral processes over a circular measure such as time-of-day. The label chronobiologists give to a particular rhythmic process references its temporal domain.

Thus, periods of less than a day are referred to as ultradian rhythms, of approximately a day as circadian rhythms and more than a day as infradian rhythms.

One of the principal aims in resolving the periodic characteristics of many biological/behavioral processes in chronobiological research is attempting to identify a single rhythm of a specific period, size and shape for a time series. In this case, the time series can be reduced and described by a small number of parameters. Typically, the procedures used in the study of temporal periodic phenomena are largely efforts in periodic regression, where a linear variable (i.e., body temperature) is regressed on a circular variable (i.e., time-of-day).

The major parameters of a periodic time series are the period or length of the cycle, amplitude (the range between the minimum and maximum of the linear variable), phase (position of the rhythm in relation to a time standard), and acrophase (the point on the circular scale where the linear variable is at a maximum. The acrophase is the phase angle. The angle measure derives from the origins of Fourier analysis. The length of the period of the rhythm is considered to be 360 (2π radians). The acrophase indexes the distance in circumference (in degrees) from a reference point (i.e., time standard). Finally, the mesor of the rhythm represents the level of the curve which indexes the origin of the circular function, or the 'center line' in the linear plot of the circular function.

Minnesota cosinor technique. One particular application of determining the periodicity in simulation training data might be found in studying the relationship between some simulation based linear performance measure and routinely scheduled subcaliber live fire training exercises. A possible question in this case might be asked on how the timing of the subcaliber exercises affect simulation performance. That is, given that subcaliber exercises are conducted quarterly, is there a particular sinusoidal rhythm in simulator performance that may be attributed to these exercises. A particular procedure for estimating the parameters described above is using the Minnesota cosiner approach, which is one of several single sinusoid fitting procedures (Halberg et al., 1977). The time series data can be generated from a longitudinal study where a small sample of individuals are monitored over a long period of time. In contrast, the series data can be generated from a transverse study where a large number of individuals each contribute a small time series. In any case, the initial stage of the analysis is to create a number of short time series by either chopping up a long series or considering each of the short series (Monk, 1981).

One begins the analysis by fitting a single sinusoid to each of the k time series (see Batschelet, 1981). The estimates of the phase and amplitude parameters are computed for each of the k series yielding k pairs of parameters $(\hat{R}_h, \hat{P}_h, h = 1, 2, 3......k)$. The parameters are then transformed to rectangular coordinates, producing k pairs of x and y values.

The analysis assumes the x and y values have been sampled from a bivariate normal distribution, thus, estimates of mean X and mean Y and the confidence interval in the X and Y dimensions can be computed. The mean is represented by the average of the X values and the average of the Y values. Once the average X and Y values are computed, one can then backtransform these values to polar coordinates and produce the final estimates of phase and amplitude:

$$\hat{R} = (X^2 + Y^2)^{1/2}$$
$$\hat{P} = \arctan(\overline{Y}/\overline{X}) \qquad (53.0)$$

As Halberg et al. (1977) indicates, valid estimates presume the k short times series to be true estimates of the final time series. Non-stationarity with respect to phase will reduce the average value of amplitude thus producing an underestimate of this parameter. Therefore, in comparing two time series one cannot discuss the issue of differences in amplitude without the assumption that both series are stationary.

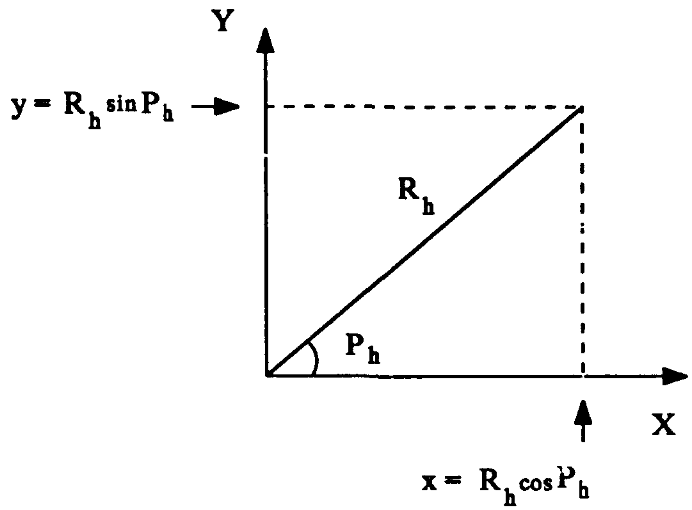

Figure 14. Transforming the polar coordinates P_h and R_h to rectangular coordinates.

The confidence limits around the mean X and Y values of the bivariate normal distribution provides for a 'significance area'. This area is equal to an ellipse on the polar plot with the point representing the final estimates of phase and amplitude at its center. The computational techniques for generating confidence ellipses are similar to those for the grand mean and can be found in Tong (1977) or Batschelet (1981). The rhythm is statistically reliable if the ellipse does not contain the origin of the coordinate system. Since non-stationarity reduces the final estimate of amplitude, then weaknesses in the assumptions of stationarity will affect the significance test (i.e., the test is sensitive to both phase and amplitude information).

Circular/linear regression. If we have information on the periodicity characteristics of the data (i.e., phase, amplitude etc.) we can apply traditional least-squares regression for fitting a regression line to the series. For example, consider simulation data on 'call-for-fire' performance. The interest here might be in examining the predictability between directional errors in call-for-fire (i.e., dependent variable) and distance from target (i.e., independent variable). One may suspect a particular bias emerges as a function of distance.

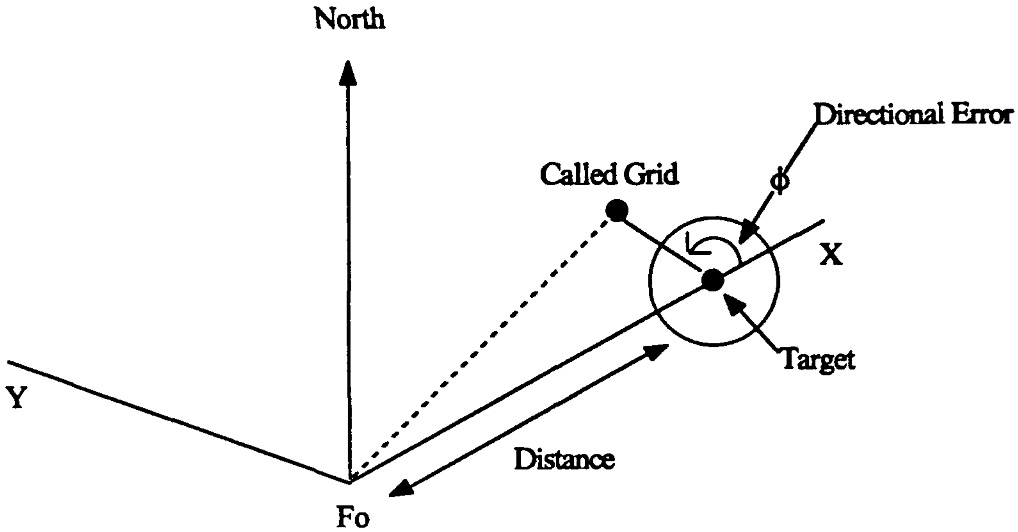

Figure 15. Representation of directional response errors in call-for-fire as a function of distance from target.

A general sinusoidal model can be defined for fitting periodic data,

$$Y = M + A\cos(\omega X - \phi) \qquad (54.0)$$

where, M is the mesor, A is amplitude, ω is angular frequency (360°/period) and ϕ is equal to the acrophase. Here, a linear dependent variable would be regressed on a circular independent variable. However, one could develop the equation for the call-for-fire example in which regression of a circular dependent variable on a linear independent variable was the object of analysis,

$$A(\sin(\omega X - \phi)) = a + b(X) \qquad (55.0)$$

where, a is a constant and b is a linear slope coefficient. In both cases, we assume that the angular frequency is given, and thus, focus on estimating the linear parameters M, b, X and Y. The least squares principle can be used to fit the data.

One might begin the call-for-fire analysis by plotting angular deviations between target position and 'called-grid' in an effort to get a feel for the distributional characteristics of the data. Clearly, if the deviation scores plot as a linear function of distance from target, traditional regression methods will apply. However, a cosine model may prove more accurate in capturing the variation in deviation scores for linear analysis. Therefore, a heuristic approach for applying circular/linear regression to the deviation call-for-fire example might be developed. First, it would be useful to generate a linear plot of the deviation scores in an effort to center the data. This allows for a clear idea as to whether a symmetric unimodal distribution is at hand, and provides one with minimum/maximum

values which illuminate phase and amplitude features of the data. Second, it is useful to compare a linear plot of the deviation scores with cosine deviation scores in order to determine which values provide the best linear fit. In addition, alternate transformations (i.e., log, square root etc.) may be found to better suit fitting the data. The caution here is that one not simply select a cosine model because of the circular nature of the problem, rather, plot the data and consider a transformation that will optimize model fit.

Linear/circular correlation. Frequently, an index characterizing the strength in the relationship between a linear and circular variable is of interest. For example, the relation between a linear performance measure such as number of hits/misses and the circular measure time-of-day may be of interest when conducting an extended or sustained simulation exercise in an effort to capture the effects of fatigue and sleep loss on simulation performance. In this case, we are not interested in estimating the parameters associated with the periodicity in the data (i.e., phase, amplitude, mesor). Instead, we are interested in determining if Y is correlated with ϕ.

Because we have assumed the period to be some value (say 360°), we can apply traditional linear zero-order correlation to the Y, ϕ data pairs.

$$\text{If: } r_{yc} = \text{corr}(Y, \cos\phi), \quad r_{ys} = \text{corr}(Y, \sin\phi), \quad r_{cs} = \text{corr}(\cos\phi, \sin\phi)$$

then we can define the correlation between the data pairs to be

$$r^2 = (r_{yc}^2 + r_{ys}^2 - 2r_{yc}\,r_{ys}\,r_{cs})/(1 - r_{cs}^2). \qquad (56.0)$$

Batschelet (1981) notes that if Y and ϕ are independent, then (nr^2) is distributed like a χ_2^2 variable, thus allowing χ^2 as the significance test.

Discussion

The results of this survey appear to indicate that traditional 'linear-based' statistics may not satisfactorily characterize the statistical properties of data generated from circular scaled variables. Further, under some conditions, these traditional methods can produce misleading information regarding the moments of a data distribution. In analyses of data taken from networked simulator tactical exercises, the usual arithmetic mean and standard deviation were shown to be inappropriate descriptive measures for circular data. The arithmetic mean angle, for example, may indicate an entirely erroneous average direction, and thus mislead a researcher on the expected value in a given data distribution. In

addition, the usual standard deviation may often be based on deviations from an inaccurate mean.

Similarly, traditional statistical analyses can not always be counted on to maximize the information contained within spatial and temporal performance measures. For example, the use of the average absolute deviations to measure directional errors of judgment discard important information on directional bias. The average direction measured by the circular mean angle provides an indication of bias in directional judgments or other directional errors. The information on directional bias may be important in developing recommendations regarding tactical system configurations or operation doctrine.

Many of the problems associated with the use of traditional statistical methods for describing circular data also emerge when discussing statistical inference. The usual parametric or nonparametric methods of statistical inference do not take into account scale circularity when it exists. Therefore, these methods will be subject to serious, often unknown and unrecognized errors in stated probabilities associated with Type 1 error rates, loss of statistical power, or both. This survey supports the notion that statistical methods are available for use that minimize the interpretational risks associated with circular data analysis when certain distributional assumptions are met.

Finally, failure to recognize the circularity of one or more variables in time series, regression, or correlation analysis may lead to overlooking important systematic relationships among variables. The use of appropriate circular methods can assist the researcher in simplifying statistical relationships and improve the fit of data models.

References

Batschelet, E. (1965). Statistical Methods For the Analysis of Problems in Animal Orientation and Certain Biological Rhythms. American Institute of Biological Sciences, Washington, D. C.

Batschelet, E. (1972). Recent statistical methods for orientation data. In: S. R. Galler, K. Schmidt-Koenig, G. J. Jacobs, and R. E. Belleview (eds.), Animal Orientation and Navigation, U. S. Government Printing Office, Washington, D. C.

Batschelet, E. (1981). Circular Statistics in Biology. New York: Academic Press.

Bingham, M. S. (1978). A characterization of the uniform distribution of the circle in the analysis of directional data. Journal of Applied Probability, 15, 852-857.

Du Bois, R. S., & Smith, P. G. (1989). A Simulation-Based Evaluation of a Position Navigation System for Armor: Soldier Performance, Training and Functional Requirements (ARI Technical Report 834). Army Research Institute For The Behavioral Sciences. (AD A210 696)

Durrand, D., & Greenwood, J. A. (1958). Modifications of the Rayleigh test for the uniformity in analysis of two-dimensional orientation data. Journal of Geology, 66, 229-238.

Fisicaro, S. A. (1989). An Evaluation of Four Alternative Orientation Indicators to Accompany the Tank Commander's Independent Thermal Viewer (ARI Technical Report 854). Army Research Institute For The Behavioral Sciences. (AD A212 788)

Halberg, F., Carandente, F., Cornelissen, G., & Katinas, G. S. (1977). Glossary of chronobiology. Chronobiologia, Supplemental 1.

Mardia, K. V. (1972). Statistics of Directional Data. New York: Academic Press.

Monk, T. H. (1981). Research methods of chronobiology. In: W. B. Webb (ed), Biological Rhythms, Sleep, and Performance. New York: John Wiley.

Taylor, D. I' , & Auburn, J. S. (1978). Orientation of amphibians by linearly polarized light. In: K Schmidt-Koenig, & W. T. Keeton (eds), Animal Migration, Navigation, and Homing. New York: Springer Verlag.

Tong, Y. L. (1976). Parameter estimation in studying circadian rhythms. Biometrics, 32, 85-94.

Watson, G. S., & Williams, E. J. (1956). On the construction of significance tests on a circle and the sphere. Biometrika, 43, 344-352.

White, S., McMeel, K., & Gross, W. (1990). Unit Performance Measurement and Training Program Designed For Networked Simulators. Technical Report No. PTR-4013-1-90, Perceptronics.

Wickens, C. D. (1984). Engineering Psychology and Human Performance. Colombus, Ohio: Charles Merrill.

Zar, J. H. (1974). Biostatistical Analysis. Englewood Cliffs, New Jersey: Prentice Hall.